Estimating the hydraulic conductivity at the South Oyster Site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model

Jinsong Chen

Department of Civil and Environmental Engineering, University of California, Berkeley, California

Susan Hubbard

Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California

Yoram Rubin

Department of Civil and Environmental Engineering, University of California, Berkeley, California

Abstract. This study explores the use of ground penetrating radar (GPR) tomographic velocity, GPR tomographic attenuation, and seismic tomographic velocity for hydraulic conductivity estimation at the South Oyster Site, using a Bayesian framework. Since site-specific relations between hydraulic conductivity and geophysical properties are often nonlinear and subject to a large degree of uncertainty such as at this site, we developed a normal linear regression model that allows exploring these relationships systematically. Although the log-conductivity displays a small variation ($\sigma^2 = 0.30$) and the geophysical data vary over only a small range, results indicate that the geophysical data improve the estimates of the hydraulic conductivity. The improvement is the most significant where prior information is limited. Among the geophysical data, GPR and seismic velocity are more useful than GPR attenuation.

1. Introduction

Heterogeneity of hydraulic conductivity in porous media controls groundwater flow and contaminant transport [*Dagan*, 1982; *Gelhar and Axness*, 1983]. Modeling of this heterogeneity is difficult and subject to a large degree of uncertainty due to the lack of densely sampled in situ hydrological measurements.

Conventional borehole techniques such as flowmeter and slug tests for collecting hydrological data are costly, time consuming, and invasive; therefore a large effort has been undertaken to explore the potential of using geophysical data to compensate for the scarcity of in situ hydrological measurements [*Rubin et al.*, 1992; *Copty et al.*, 1993; *Copty and Rubin*, 1995; *Hubbard et al.*, 1997; *Rubin et al.*, 1998; *Ezzedine et al.*, 1999; *Hubbard and Rubin*, 2000]. Geophysical data used for hydrogeological characterization often include electrical resistivity [*Kelly*, 1977; *Ahmed et al.*, 1988], seismic velocity [*Rubin et al.*, 1992; *Copty et al.*, 1993; *Copty and Rubin*, 1995; *Hyndman et al.*, 1994], and ground penetrating radar (GPR) velocity [*Hubbard et al.*, 1997, 1999]. Methods for integration of hydrological and geophysical data include regression models [*Kelly*, 1977], cokriging models [*Ahmed et al.*, 1988], inversion models [*Rubin et al.*, 1992], and Bayesian models [*Copty et al.*, 1993; *Ezzedine et al.*, 1999]. Despite the difference in the methods and the geophysical data, it has been widely recognized that the most difficult part of the integration is tying hydrological measurements to geophysical data because of the scale and

resolution disparity between hydrological and geophysical measurements [*Ezzedine et al.*, 1999] and because of their non-unique relationships due to the uncertainty associated with field data acquisition and interpretation [*Urish*, 1981]. This paper proposes an approach to dealing with this issue based on the normal linear regression model. It extends the previous work reported by *Copty et al.* [1993] and *Ezzedine et al.* [1999], and its main novelty is in formulating the petrophysical models in a probabilistic fashion, using likelihood functions.

The paper explores the potential use of GPR tomographic velocity, GPR tomographic attenuation, and seismic tomographic velocity as well as hydrological measurements for estimating hydraulic conductivity. It is focused on the usefulness of geophysical measurements for hydraulic conductivity estimation and on the integration of hydrological and geophysical data.

Our approach in this paper is to explore the correlations between the geophysical attributes and the hydraulic conductivity, following the ideas explored by *Rubin et al.* [1992], *Copty et al.* [1993], *Hyndman et al.* [1994], and *Ezzedine et al.* [1999]. Another approach to this problem is to analyze transient effects, for example, through time-lapse tomography [*Shapiro et al.*, 1999]. This approach was not pursued here because it requires injecting fluids, which can potentially influence the geophysical signals and thus may have detrimental effects on the overall goals of our field experiments.

The remaining of this paper is organized as follows. Section 2 introduces the South Oyster Site, available data at the site, and some preliminary data analyses. Section 3 describes the Bayesian method and the normal linear regression model. Section 4 explores the use of geophysical data within a Bayesian
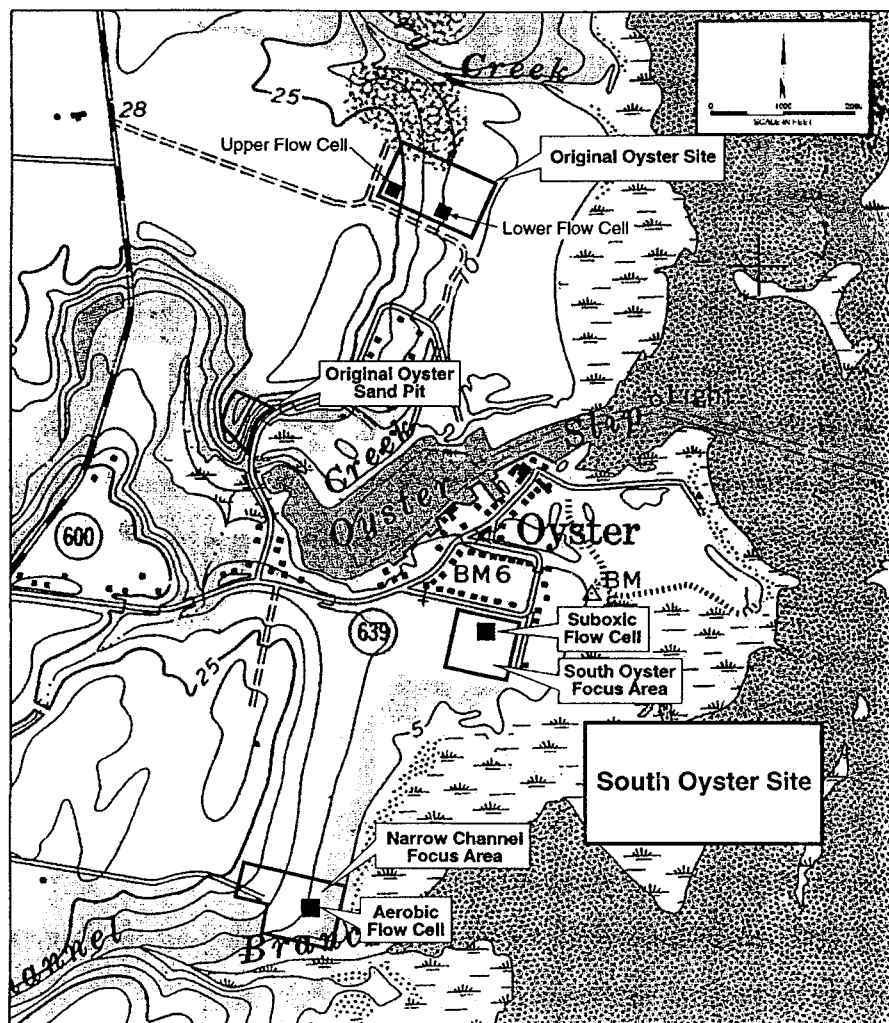
Figure 1. Locations of the South Oyster Site and the Aerobic Flow Cell.

framework for estimating hydraulic conductivity. Discussion of our findings and conclusions are given in section 5.

2. Site and Data Descriptions

2.1. South Oyster Site

The South Oyster Site is located near the town of Oyster on Virginia's Eastern Shore Peninsula between the Chesapeake Bay and the Atlantic Ocean. A field-scale experiment has been undertaken by a multidisciplinary research team within an uncontaminated aquifer at the Oyster Site to evaluate the importance of chemical and physical heterogeneity in controlling bacteria that are injected into the ground for bioremediation purposes (M. F. DeFlaun et al., Transport of bacteria in a coastal plain aquifer under forced gradient conditions, submitted to *Eos Transactions AGU*, 2000]. The sediments at the South Oyster Site consist of unconsolidated to weakly indurated, well-sorted, medium- to fine-grained Late Pleistocene sands and pebbly sands. The upper 9 m of the South Oyster Site consists of the Wachapreague Formation, which was deposited in a shallow, open marine to back-barrier environment, north of the tide-dominated mouth of the Chesapeake Bay [*Mixon*, 1985]. The water table at the South Oyster Site is located ~3 m below ground surface.

Within the South Oyster Site, two study focus areas exist: the South Oyster Focus Area and the Narrow Channel Focus Area (Figure 1). Locations of the focus areas were chosen based primarily on groundwater chemistry: the South Oyster Focus Area is situated within a suboxic portion of the aquifer, while conditions at the Narrow Channel Focus Area are predominantly aerobic. Forced gradient chemical and bacterial tracer test experiments were performed within what is called the "Aerobic Flow Cell" of the Narrow Channel Focus Area in 1999 [*Johnson et al.*, 2001] and within the "Suboxic Flow Cell" of the South Oyster Focus Area in 2000. At both locations, extensive geological, geophysical, and hydrological data have been collected to characterize the subsurface prior to the tracer test experiments. In this study, we explore the use of geophysical tomographic data, collected within the saturated portion of the Aerobic Flow Cell (approximately between depths of 0–6.0 m below mean sea level (msl)), for providing detailed hydraulic conductivity estimates there.

The Aerobic Flow Cell layout within the Narrow Channel Focus Area is shown in Figure 2. Hydraulic conductivity measurements are available at the well bore locations, indicated by circles, and geophysical tomographic profiles are available between several well pairs, as indicated by the solid lines. De-
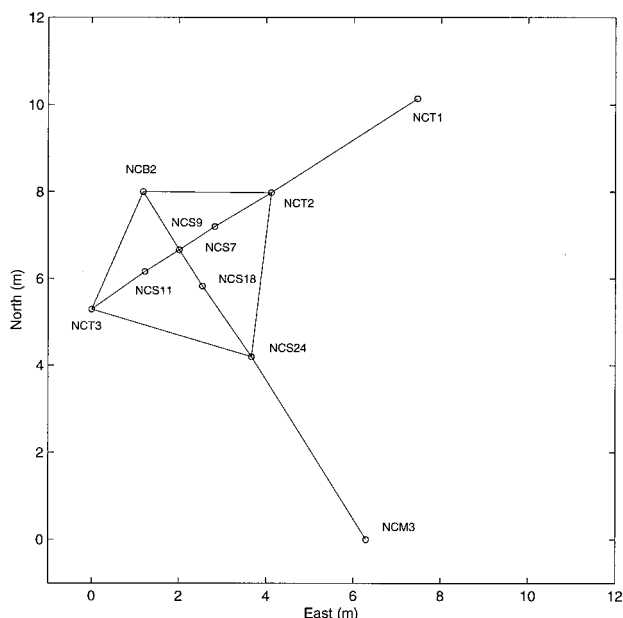
Figure 2. Aerobic Flow Cell in the Narrow Channel Focus Area (the circles denote flowmeter measurement well bores and the solid lines denote geophysical tomographic profiles).

scriptions of these available data are discussed in sections 2.2.1 and 2.2.2, respectively. The chemical and bacterial tracer injection well is NCB2. Groundwater flow direction and geologic dip are aligned approximately parallel to the transect NCB2-NCM3, and geologic strike is aligned approximately parallel with the transect NCT3-NCT1. Twenty-four multilevel samplers were installed between the wells NCB2 and NCM3 and NCT3 and NCT2 to detect the passage of chemical and bacterial tracers over time during the tracer test experiments as described by *Johnson et al.* [2001]. The log-conductivity estimates within the Aerobic Flow Cell, obtained using geophysical tomographic data as described in this study, will be used to help constrain the stochastic numerical flow studies being performed to understand the transport experiment results [*Scheibe et al.*, 1999].

2.2. Field Sampling

2.2.1. Hydraulic conductivity measurements from flowmeter data.
Hydraulic conductivity values were calculated from flowmeter and slug test data within the Aerobic Flow Cell. Electromagnetic borehole flowmeter data were collected from all wells whose locations are shown in Figure 2. Each well was ~9.4 m deep and had two 3.05 m long screens positioned in the lower 6.1 m, or from ~0.5 m to 5.8 m below msl [*Waldrop and Hamby*, 1999]. The flowmeter data provided relative hydraulic conductivity measurements at discrete intervals of 0.15 m for each well. Slug test data, where available, were used to provide average hydraulic conductivity values over the screened well intervals. Where slug test data were not available, the geometric mean of the slug tests over the entire flow cell was used. These average hydraulic conductivity values were then used to convert the relative hydraulic conductivity measurements, obtained from flowmeter data, into absolute hydraulic conductivity values for that well [*Molz and Young*, 1993].

2.2.2. Geophysical measurements from tomographic GPR and seismic data.
Both GPR and seismic tomographic data were collected within the Aerobic Flow Cell along the transects indicated in Figure 2. For tomographic acquisition geometry, GPR transmitting antenna (or seismic source) and GPR receiving antennas (or seismic geophones) are located in separate well bores, and direct energy from a transmitting antenna in one well bore is recorded by a receiving antenna located in the other well bore. The transmitter position is changed and the recording repeated until both the transmitter and the receiver have occupied all possible positions within the two well bores. The direct electromagnetic wave or seismic *P* wave travel time between all transmitter/receiver positions, as well as the amplitude of the direct arrival, is obtained from the recorded data. The interwell area is then discretized into a grid composed of cells or pixels, and inversion algorithms are used to transform the recorded travel time and amplitude information into estimates of velocity and attenuation, respectively, at each pixel. The discretization that is chosen for the inversion is typically based on consideration of several factors including: the wavelength of the propagating signal, expected material properties and their contrasts, acquisition geometry including well bore spacing and geophone spacing (which control propagation distance, ray path density, and illumination angles), and inversion damping parameters. The reader interested in different types of geophysical inversion approaches and applications is referred to *Parker* [1994], *Williamson and Worthington* [1993], and *Rector* [1995].

The seismic tomographic profiles were collected along the same transects where GPR tomographic data were collected (Figure 2). These data were collected using a Geometrics Strataview seismic system with a piezoelectric source. The central frequency of the pulse was 4000 Hz, with a bandwidth from ~1000 to 7000 Hz, rendering an average seismic wavelength of ~0.4 m. The source and geophone spacing in the well bores was 0.125 m, which resulted in dense ray path coverage of over 13,200 traces in the interwell area. On the basis of the seismic wavelength, small well bore spacing of ~3–5 m, and dense ray path coverage over a variety of illumination angles, a discretization of 0.25 m × 0.25 m was chosen for the discretization inversion. The travel times were picked for all source-receiver pairs. The travel time data were then inverted using a straight-ray algebraic reconstruction technique [*Peterson et al.*, 1985] to produce seismic velocity estimates for each 0.25 m × 0.25 m cellblock along all transects. The small velocity range observed in the data suggests that distortion caused from ray bending should be minimal and that ray path density should be fairly evenly distributed in the interwell area.

Seismic amplitudes can yield information about the attenuative properties of subsurface sediments or rocks. In a manner similar to travel time inversion, amplitude information can be extracted from the tomographic data and inverted for attenuation in the interwell area. Many theories exist to incorporate the great variety of mechanisms that can influence seismic wave attenuation. For porous, granular, sedimentary rocks, the generally accepted mechanisms may be grouped into three broad categories: scattering attenuation, fluid-flow attenuation, and fluid-matrix attenuation [*Nihei*, 1992]. Because of the variety of influences on the seismic amplitudes, it is often difficult to extract meaningful characterization information from seismic amplitude data. Additionally, seismic amplitudes are extremely sensitive to the presence of trapped gas and the state of consolidation. The presence of a shallow water table,

small amounts of organic material (potential sources of trapped gas), and varying states of consolidation of the Oyster sediments are suspected to have affected the seismic source radiation pattern and coupling, as well as the receiver properties. If not adjusted, these radiation pattern and coupling variations often yield inversion artifacts [*Vasco et al.*, 1996; *Keers et al.*, 2000]. Because inversion artifacts were observed in the Oyster seismic attenuation tomograms, our confidence in these data was not high enough to allow us to use them in our hydrological property estimation procedure. Use of the Oyster seismic amplitude data is left for future studies, after further investigation and preprocessing of the seismic radiation pattern and coupling effects have been performed.

The tomographic GPR data were collected using a PulseEKKO 100 system with 200 MHz central frequency well bore antennas. These data were collected using a transmitter/receiver spacing in the well bores of 0.125 m, which again resulted in over 13,200 traces per tomogram. The effective range of the radar propagation frequencies was 40–140 MHz, rendering effective GPR wavelengths of ~0.5 m. Although the wavelengths of the radar data are on average greater than those of the seismic data, the high signal-to-noise ratio of the radar data relative to the Oyster seismic data permitted inversion using the same discretization as was used for the seismic data of 0.25 m $\times$ 0.25 m. The similar discretization used for both the seismic and radar tomographic data inversions facilitated the hydrogeological parameter estimate computations. For the high radar frequencies employed and in the sandy environment at the Narrow Channel Focus Area, the radar propagation velocities are primarily governed by variations in the dielectric constant, and the amplitudes are primarily affected by variations in dielectric constant and electrical conductivity [*Davis and Annan*, 1989] of the interwell sediments. Unlike the seismic amplitude data, the Oyster radar amplitude radiation patterns and source-receiver coupling appeared to be consistent, and thus inversion was performed on both the picked travel times and amplitudes using straight-ray algebraic reconstruction techniques [*Peterson et al.*, 1985; *Peterson*, 2001] to yield electromagnetic wave velocity and attenuation estimates for each cellblock along all tomographic profiles.

2.3. Data Analysis

Since our plan is to explore and test the use of geophysical tomographic measurements for hydraulic conductivity estimation, we employ geophysical data only at the well bores where hydraulic conductivity data are available. Geophysical data at these wells were approximated by the ones at the cellblocks located one column away from the wells rather than directly near the wells on their surrounding transects. This is because geophysical data located directly near the wells may be affected by the presence of disturbed zones around the wells [*Peterson*, 2001]. For some wells such as NCB2, there are three transects passing them; we averaged the geophysical data extracted from each transect. Consequently, we can obtain geophysical data at each well location with an interval of 0.25 m along vertical directions. Since log conductivity was sampled with an interval of 0.15 m rather than 0.25 m, we interpolated the log conductivity data at vertical locations where geophysical data were also sampled for each well. Finally, we can create a data set in which each log conductivity value has corresponding collocated geophysical data.

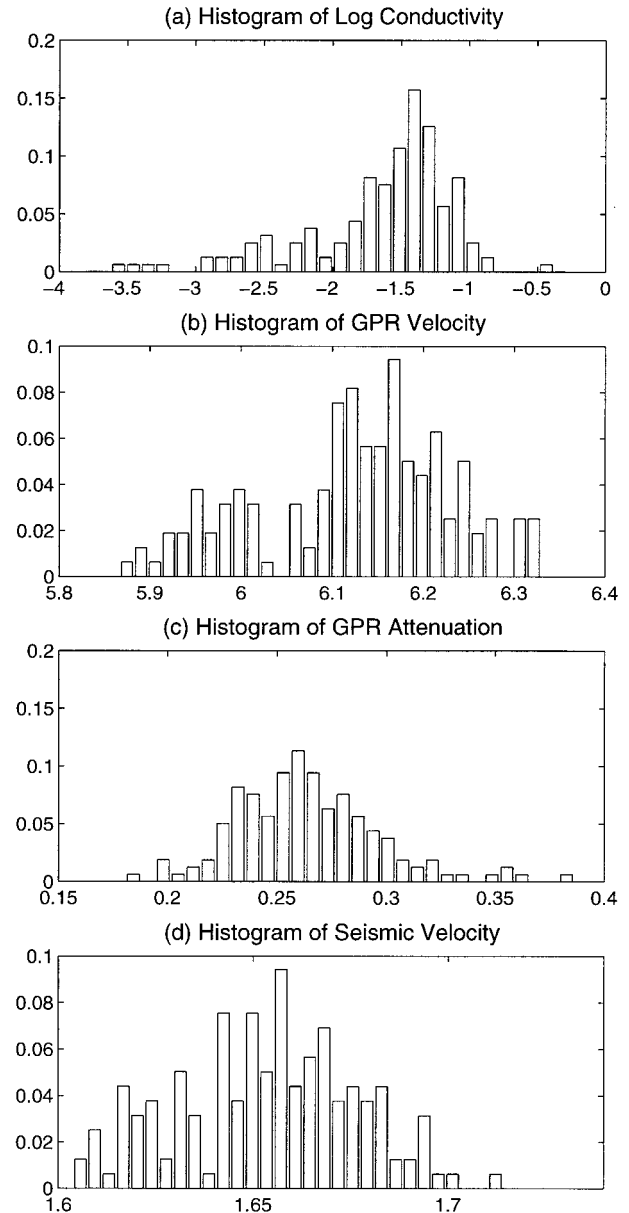**2.3.1. Hydraulic conductivity.** The histogram of natural log conductivity at the Aerobic Flow Cell is shown in Figure 3a.

Figure 3. (a) Histogram of log-conductivity (hydraulic conductivity in m h^{-1}). (b) Histogram of GPR velocity (cm ns^{-1}). (c) Histogram of GPR attenuation (1 m^{-1}). (d) Histogram of seismic velocity (km s^{-1}).

It is asymmetric and negatively skewed and suggests the existence of two subpopulations of hydraulic conductivity, high- and low-conductivity zones [*Copty et al.*, 1995; *Welhan and Reed*, 1997].

The spatial structure of the log conductivity was identified through covariance analyses, and the results are shown in Figure 4. The vertical covariance given in Figure 4a can be fitted with an exponential covariance model with a range of 0.6 m and a sill of 0.30 for small lags (<0.6 m), but hole-type structure appears at large lags (>0.6 m). The hole effect may be the outcome of repetitive sequences or periodic variations [*Isaaks and Srivastava*, 1989]. As pointed out by *Journel and Huijbregts* [1978], the hole effect may also be due to an artificial pseudo-periodicity of available data and can be ignored in practice if not very remarkable. The lateral covariances along the direc-
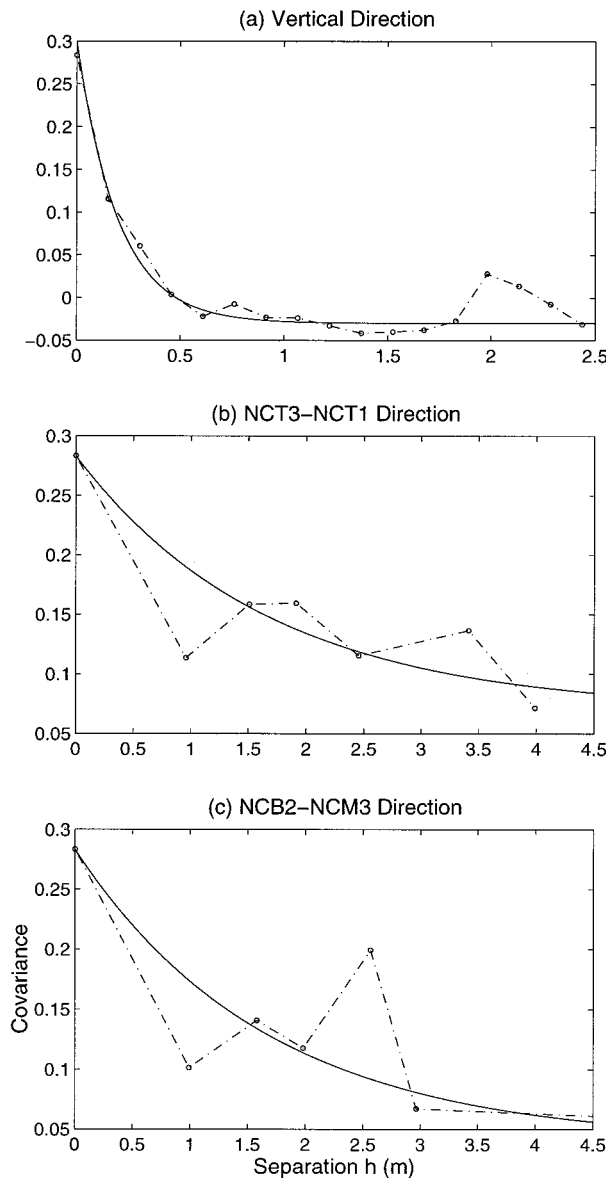
Figure 4. (a) Covariance along vertical direction. (b) Covariance along the direction perpendicular to geologic strike (NCB2-NCM3). (c) Covariance along the direction parallel to geologic strike (NCT3-NCT1).

tions perpendicular and parallel to the geologic strike are shown in Figure 4b and Figure 4c, respectively. Both covariances are fitted with exponential models with a range of 5 m.

2.3.2. Geophysical data. Histograms of GPR velocity, GPR attenuation, and seismic velocity are shown in Figures 3b–3d, respectively. Similar to log conductivity (Figure 3a), negative skewness is observed in the histograms of GPR and seismic velocity. This suggests correlations between log conductivity and GPR and seismic velocity. The GPR and seismic velocity change over small ranges and exhibit small variations (coefficient of variation CV = 1.7% for GPR velocity and CV = 1.4% for seismic velocity) compared to the GPR attenuation (CV = 12.5%).

2.3.3. Correlations between log conductivity and geophysical data. Physical connections between log conductivity and GPR velocity, GPR attenuation and seismic velocity exist but
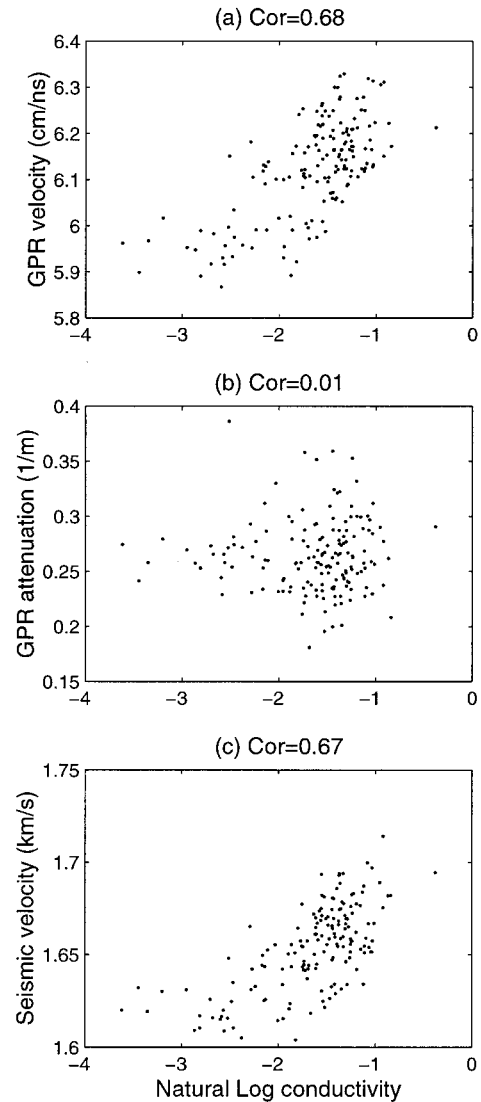

Figure 5. Scatterplots of log conductivity, GPR velocity, GPR attenuation, and seismic velocity.

are not straightforward. For instance, hydraulic conductivity correlates to porosity, as evidenced by the Kozeny-Carman equation [*Carman*, 1956], and GPR velocity, GPR attenuation, and seismic velocity also relate to porosity [*Knoll*, 1996; *Marion*, 1990; *Mavko et al.*, 1998].

Figure 5 depicts scatterplots of log conductivity versus GPR velocity, GPR attenuation, and seismic velocity based on the data available at the Aerobic Flow Cell. The log conductivity correlates with GPR and seismic velocity, and it generally increases as GPR and seismic velocity increase. The GPR attenuation and log conductivity appear to be uncorrelated, and the GPR attenuation associated with low log conductivity [$\log(k) < -2$] seems to be less variable.

3. Methodology

Estimating log conductivity suffers from much uncertainty owing to the lack of densely sampled in situ hydrological measurements and owing to the absence of unique relations between log conductivity and geophysical data. To address this uncertainty, we adopt a stochastic framework in which log

conductivity, GPR velocity, GPR attenuation, and seismic velocity are considered as spatial random functions.

3.1. Bayesian Formula

In this section, we develop a Bayesian methodology for estimating log conductivity from hydrological and geophysical data. Let the random variable Y denote log conductivity and V_g, α, and V_s denote GPR velocity, GPR attenuation, and seismic velocity, respectively. All data are mean-removed and normalized by their corresponding standard deviations. The log conductivity estimate at a given location $\mathbf{x}$, in terms of probability density function (pdf), is obtained using the Bayes theorem as follows [Box and Tiao, 1973; Kitanidis, 1986]:

$$f'_Y[y(\mathbf{x})] = CL[y(\mathbf{x})|v_g(\mathbf{x}), \alpha(\mathbf{x}), v_s(\mathbf{x})]f_Y[y(\mathbf{x})], \quad (1)$$

where $y(\mathbf{x})$ is an unknown value of Y being estimated at $\mathbf{x}$, $v_g(\mathbf{x})$, $\alpha(\mathbf{x})$, and $v_s(\mathbf{x})$ are the measured values of V_g, α, and V_s at the same location, C is a normalizing coefficient, $L[y(\mathbf{x})|v_g(\mathbf{x}), \alpha(\mathbf{x}), v_s(\mathbf{x})]$ is the likelihood function given $v_g(\mathbf{x})$, $\alpha(\mathbf{x})$, and $v_s(\mathbf{x})$, and $f'_Y[y(\mathbf{x})]$ and $f_Y[y(\mathbf{x})]$ are the posterior and prior pdfs of Y at $\mathbf{x}$, respectively. Note that only collocated geophysical data have been used to update the prior pdf since they are most informative compared to the measurements at adjacent locations [Copty et al., 1993].

The Bayesian method has been used for many years in the water resources field. One of the earliest applications in groundwater hydrology was provided by Kitanidis [1986] for analyzing parameter uncertainty in estimation of spatial functions. In that work, the mean and covariance matrix of the posterior distribution were derived analytically by choosing a prior distribution that is conjugate to the likelihood function in the sense that the posterior has the same form as the prior. Following the same line, Copty et al. [1993] applied the method to subsurface characterization of hydrological properties using geophysical data, and the analytical forms of the posterior mean and variance were also obtained under certain assumptions. In this study, we will develop a new approach, which allows for large flexibility in the form of the likelihood function and posterior pdf, to get numerical rather than analytical posterior mean and variance.

3.1.1. Prior pdf. The prior pdf $f_Y[y(\mathbf{x})]$ was estimated based on the hydraulic conductivity data using kriging [Journel, 1989]. A similar approach was also used by Copty et al. [1993] and Ezzedine et al. [1999]. The prior distribution is normal if Y is multivariate normal [Deutsch and Journel, 1998].

3.1.2. Likelihood function. The likelihood function $L[y(\mathbf{x})|v_g(\mathbf{x}), \alpha(\mathbf{x}), v_s(\mathbf{x})]$ plays a central role in the Bayesian method and was inferred from the hydrological and collocated geophysical data. It is expressed as follows [Bernardo and Smith, 1994]:

$$L[y(\mathbf{x})|v_g(\mathbf{x}), \alpha(\mathbf{x}), v_s(\mathbf{x})]$$
$$= f_{V_g}[v_g(\mathbf{x})|y(\mathbf{x})]f_\alpha[\alpha(\mathbf{x})|y(\mathbf{x}), v_g(\mathbf{x})]$$
$$\cdot f_{V_s}[v_s(\mathbf{x})|y(\mathbf{x}), v_g(\mathbf{x}), \alpha(\mathbf{x})], \quad (2)$$

where $f(\ |\ )$ denotes a conditional pdf. If V_g, α, and V_s are independent such that $f_\alpha[\alpha(\mathbf{x})|y(\mathbf{x}), v_g(\mathbf{x})] = f_\alpha[\alpha(\mathbf{x})|y(\mathbf{x})]$ and $f_{V_s}[v_s(\mathbf{x})|y(\mathbf{x}), v_g(\mathbf{x}), \alpha(\mathbf{x})] = f_{V_s}[v_s(\mathbf{x})|y(\mathbf{x})]$, the inference of the likelihood function becomes simple since each conditional pdf involves only two variables. This is, however, not the case in the present study where there are four depen-

dent variables (Y, V_g, α, and V_s), and we need to update the prior pdf based on all the collocated geophysical data.

3.2. Normal Linear Regression Model

The normal linear regression model [Stone, 1995] provides a systematic approach to the inference of the conditional pdfs shown in (2). A similar approach has also been suggested by Kitanidis [1991] to model a linear drift of a spatially dependent variable, such as log conductivity. In this section, we will demonstrate the inference of $f_{V_s}[v_s(\mathbf{x})|y(\mathbf{x}), v_g(\mathbf{x}), \alpha(\mathbf{x})]$, and the method can be applied to the other functions appearing in (2).

In the normal linear regression model, the seismic velocity V_s at $\mathbf{x}$ is assumed to be normally distributed with mean μ and variance σ^2. The mean μ is assumed to be a member of the linear function space G whose basis functions consist of m distinct monomials $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $\cdots$, $g_m(\mathbf{x})$, formed from combinations of powers and products of $y(\mathbf{x})$, $v_g(\mathbf{x})$, and $\alpha(\mathbf{x})$, such as 1, $y(\mathbf{x})$, $v_g(\mathbf{x})$, $\alpha(\mathbf{x})$, $y^2(\mathbf{x})$, $v_g^2(\mathbf{x})$, $\alpha^2(\mathbf{x})$, $y(\mathbf{x})v_g(\mathbf{x})$, $y(\mathbf{x})\alpha(\mathbf{x})$, and $v_g(\mathbf{x})\alpha(\mathbf{x})$. It is modeled as follows:

$$\mu(\mathbf{x}) = \sum_{i=1}^{m} \beta_i g_i(\mathbf{x}), \quad (3)$$

where β_i is a coefficient of basis function $g_i(\mathbf{x})$, $i = 1, 2, \cdots, m$. The variance σ^2 is taken to be a constant, independent of $y(\mathbf{x})$, $v_g(\mathbf{x})$, and $\alpha(\mathbf{x})$. The final set of basis functions is determined by following a model selection procedure, given in section 3.2.2.

3.2.1. Estimation of μ and σ^2. The mean function $\mu(\mathbf{x})$ and the variance σ^2 are estimated from the data $(v_s(\mathbf{x}_j), y(\mathbf{x}_j), v_g(\mathbf{x}_j), \alpha(\mathbf{x}_j))$, $j = 1, 2, \cdots, n$, where $y(\mathbf{x}_j)$ is the log conductivity at location $\mathbf{x}_j$ and $v_s(\mathbf{x}_j)$, $v_g(\mathbf{x}_j)$, and $\alpha(\mathbf{x}_j)$ are the dimensionless collocated seismic velocity, GPR velocity, and GPR attenuation, respectively.

Estimating β_1, β_2, $\ldots$, β_m is achieved by minimizing the residual sum of squares

$$\text{RSS} = \sum_{j=1}^{n} [v_s(\mathbf{x}_j) - \mu(\mathbf{x}_j)]^2. \quad (4)$$

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ and $\mathbf{Z} = [v_s(\mathbf{x}_1), v_s(\mathbf{x}_2), \cdots, v_s(\mathbf{x}_n)]^T$, where the exponent T denotes the transpose operator. The estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, which minimizes (4), is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{Z}, \quad (5)$$

where $\mathbf{D}$ is a design matrix, given by

$$\begin{pmatrix} g_1(\mathbf{x}_1) & g_2(\mathbf{x}_1) & \cdots & g_m(\mathbf{x}_1) \\ g_1(\mathbf{x}_2) & g_2(\mathbf{x}_2) & \cdots & g_m(\mathbf{x}_2) \\ \cdots & \cdots & \cdots & \cdots \\ g_1(\mathbf{x}_n) & g_2(\mathbf{x}_n) & \cdots & g_m(\mathbf{x}_n) \end{pmatrix}. \quad (6)$$

Once β_1, β_2, $\ldots$, β_p are estimated, the mean and the variance of $f_{V_s}[v_s(\mathbf{x})|y(\mathbf{x}), v_g(\mathbf{x}), \alpha(\mathbf{x})]$ are defined as follows:

$$\hat{\mu}(\mathbf{x}) = \sum_{i=1}^{m} \hat{\beta}_i g_i(\mathbf{x}), \quad (7)$$

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{j=1}^{n} [v_s(\mathbf{x}_j) - \hat{\mu}(\mathbf{x}_j)]^2. \quad (8)$$
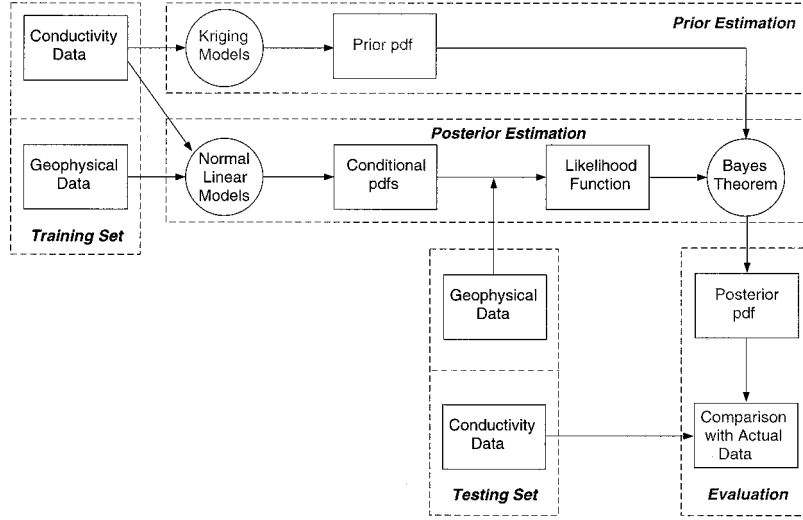
Figure 6. Schematic map of the Bayesian method.

We note that the estimate $\hat{\sigma}^2$ of the variance in (8) is stationary and independent of locations. The estimate $\hat{\mu}(\mathbf{x})$ of the mean function in (7), however, depends on $y(\mathbf{x})$, $v_g(\mathbf{x})$, and $\alpha(\mathbf{x})$ and thus on the location $\mathbf{x}$.

3.2.2. Selection of basis functions. Selecting and eliminating basis functions is the key to the normal linear regression model. We first choose all possible distinct monomials of $y(\mathbf{x})$, $v_g(\mathbf{x})$, and $\alpha(\mathbf{x})$ with a degree of 4 at most as the initial set of basis functions. By deleting some of the initial basis functions based on certain criteria described in the following, we obtain the final set.

Removing or retaining a basis function $g_i(\mathbf{x})$ $(1 \leq i \leq m)$ from the initial set is based on testing of the null hypothesis H_o: $\beta_i = 0$, at the common testing level 0.05. We first compute the statistic $u = \hat{\beta}_i / \text{SE}(\hat{\beta}_i)$, where $\text{SE}(\hat{\beta}_i)$ is the standard error of $\hat{\beta}_i$, which is the ith element in the diagonal of the matrix $\hat{\sigma}^2(\mathbf{D}^T\mathbf{D})^{-1}$ (we can obtain $\hat{\sigma}^2$ from (8) and $\mathbf{D}$ from (6), and thus the matrix $\mathbf{M} = \hat{\sigma}^2(\mathbf{D}^T\mathbf{D})^{-1}$ and $\text{SE}(\hat{\beta}_i) = \mathbf{M}(i, i)$). The statistic u follows the t distribution with $n - m$ degrees of freedom, based on the properties of the normal linear regression model [*Stone*, 1995]. We then calculate the p value, defined by $2[1 - t_{n-m}(|u|)]$, where t_{n-m} is the probability function of the t distribution with $n - m$ degrees of freedom and compare it with the testing level 0.05. If the p value is larger than 0.05, we accept the null hypothesis and remove $g_i(\mathbf{x})$ from the basis function set, but otherwise it is retained.

The procedure of deleting basis functions is an iterative process, which is executed as follows: (1) fitting a model $\mu(\mathbf{x})$ as shown in (3) to the sampled data and obtaining $\hat{\beta}_i$ (equation (5)) and $\text{SE}(\hat{\beta}_i)$ $(i = 1, 2, \cdots, m)$, (2) computing their corresponding p values, (3) comparing all the p values with 0.05, and (4) removing the basis function with the largest p value from the initial set. The procedure is repeated until no members of the basis function set can be removed. The same method is also applied to determine the other pdfs, which compose the likelihood function (equation (2)).

4. Hydraulic Conductivity Estimation

4.1. Outline of the Approach

To apply and test our approach, we split the hydraulic conductivity and geophysical data available at the Aerobic Flow Cell into a training set and a testing set. To avoid the bias in selecting the testing set, we consider each well shown in Figure 2 in turn as a testing well; thus we have 10 different combinations of training and testing sets. For each of the combination, we follow the steps outlined in Figure 6, which are grouped into the following categories:

1. First is prior estimation. We first infer the spatial correlation structure of log conductivity from the hydraulic conductivity data of the training set and then estimate the mean log conductivity and its variance at each testing location using kriging. The prior is assumed to be normally distributed with the mean and the variance as obtained from the kriging procedure.

2. Second is posterior estimation. We derive the conditional pdfs, which form the likelihood function as shown in (2), from both hydraulic conductivity and geophysical data of the training set using the normal linear regression model. We then apply the likelihood function on the prior pdf to obtain the posterior pdf by following (1).

3. Finally is evaluation. We compare various statistics of the prior and posterior pdfs with the actual measurements. More details on that are given in sections 4.2 and 4.3.

4.2. Estimating the Hydraulic Conductivity Using GPR Velocity

Figure 7 compares the log conductivity measurements at well NCS7 with the means of the prior pdfs and the posterior pdfs updated using its collocated GPR velocity only. The improvement of the posterior mean estimates is not significant, and both prior and posterior means follow quite closely the trend outlined by the actual measurements. This is because well NCS7 is close to wells NCS9 and NCT2 (Figure 2) which belong to its training set, and thus the prior mean estimates based on the hydraulic conductivity data alone are fairly accurate. Negligible improvement was also observed at other testing wells located in close proximity to training wells, such as wells NCS9, NCS11, and NCS18.

As the distances between wells of the training sets and the testing wells increase, the contribution of the collocated GPR velocity data becomes apparent. This is demonstrated in Figure 8, which compares actual measurements with the means of prior and posterior pdfs at well NCM3, located no less than
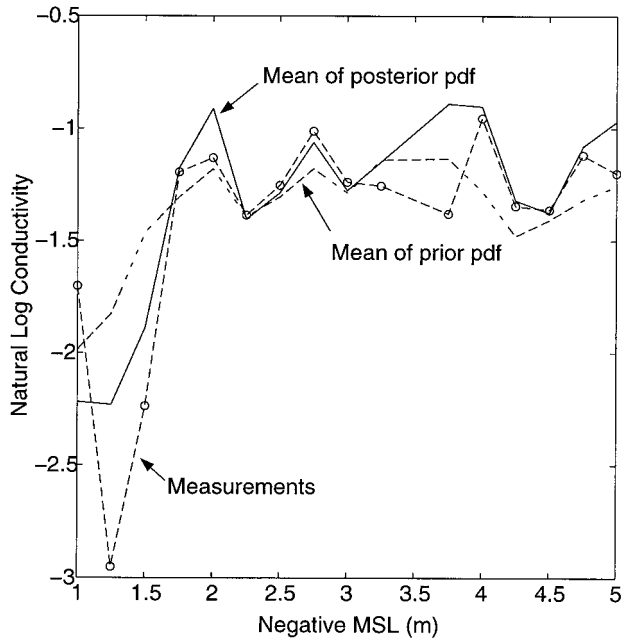
Figure 7. Comparisons of measurements, prior and posterior means at testing well NCS7.

4.95 m away from other training wells. The posterior estimates follow more closely the measurements compared to the prior estimates. Similar results were also observed at the testing wells NCT1, NCT2, NCT3, NCS24, and NCB2.

Table 1 shows the vertical averages of standard deviations of prior pdfs and posterior pdfs updated using GPR velocity for each well shown in Figure 2. The prior variances are estimated from kriging, and the posterior variances are computed from the posterior pdfs using numerical integration:
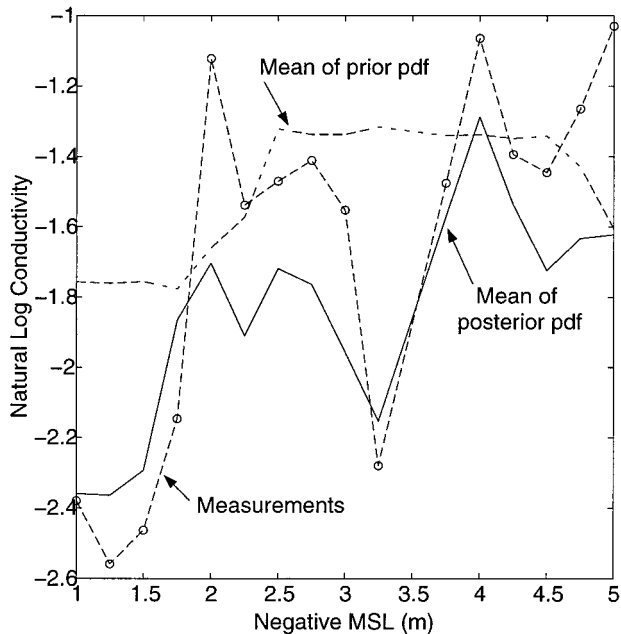

Figure 8. Comparisons of measurements and prior and posterior means at testing well NCM3.

Table 1. Comparisons of Prior and Posterior Standard Deviations

Testing Well	Prior Standard Deviations, σ_{prior}	Posterior Standard Deviations, σ_{post}	Reduction (%) $(\sigma_{prior} - \sigma_{post})/\sigma_{prior}$
NCS7	0.47	0.38	19
NCS9	0.42	0.32	24
NCS11	0.47	0.37	21
NCS18	0.49	0.37	24
NCM3	0.60	0.42	30
NCB2	0.53	0.38	28
NCS24	0.55	0.40	27
NCT1	0.59	0.42	29
NCT2	0.52	0.39	25
NCT3	0.53	0.40	25

$$\sigma_{post}^2(\mathbf{x}) = \int \left[y(\mathbf{x}) - \langle Y(\mathbf{x}) \rangle \right]^2 f'_Y[y(\mathbf{x})] \, dy, \qquad (9)$$

where $\langle Y(\mathbf{x}) \rangle$ is the mean of the posterior pdf $f'_Y[y(\mathbf{x})]$ at $\mathbf{x}$. For all the testing wells, the standard deviations are significantly reduced by the use of GPR velocity, and the reductions at wells NCM3 and NCT1 are most evident because their prior estimates are less informative owing to the relatively large distances of these wells from the wells used for calculating the priors.

Figure 9 shows the 95% confidence intervals at well NCM3 for prior and posterior estimates. The errors predicted by the prior model are on the conservative side, and at each testing location, the updating provides more narrow bounds, which are consistent with the actual errors.

4.3. Estimating the Hydraulic Conductivity Using GPR Velocity, GPR Attenuation, and Seismic Velocity

This section explores the efficiency of using GPR velocity, GPR attenuation, and seismic velocity in the hydraulic conductivity estimation. Since the addition of GPR attenuation and seismic velocity after using GPR velocity did not lead to further changes in the estimates of the mean log conductivity in our current study, our discussion focuses on prior and posterior variances of all the ten testing wells.

Figure 10 compares the averaged actual errors over the 10 testing wells with the averaged standard deviations of prior and posterior pdfs along depth and over the 10 testing wells for different combinations of geophysical data. The actual errors are the spatial averages of absolute differences between actual and estimated values along each testing well. It is evident that the standard deviations of the model using the hydraulic conductivity data only are much larger than the actual errors, but with the addition of various types of geophysical data, the standard deviations consistently decrease until they are of the same order as the actual errors.

To evaluate the efficiency of geophysical data in reducing uncertainty, we use different combinations of geophysical data in the hydraulic conductivity estimation and calculate their reductions in the standard deviations, given by

$$\frac{\sigma_{prior} - \sigma_{post}}{\sigma_{prior}} (100\%), \qquad (10)$$

where σ_{prior} and σ_{post} denote the vertical averages of the standard deviations of the prior model and of the posterior model

(a) Prior Estimates at NCM3
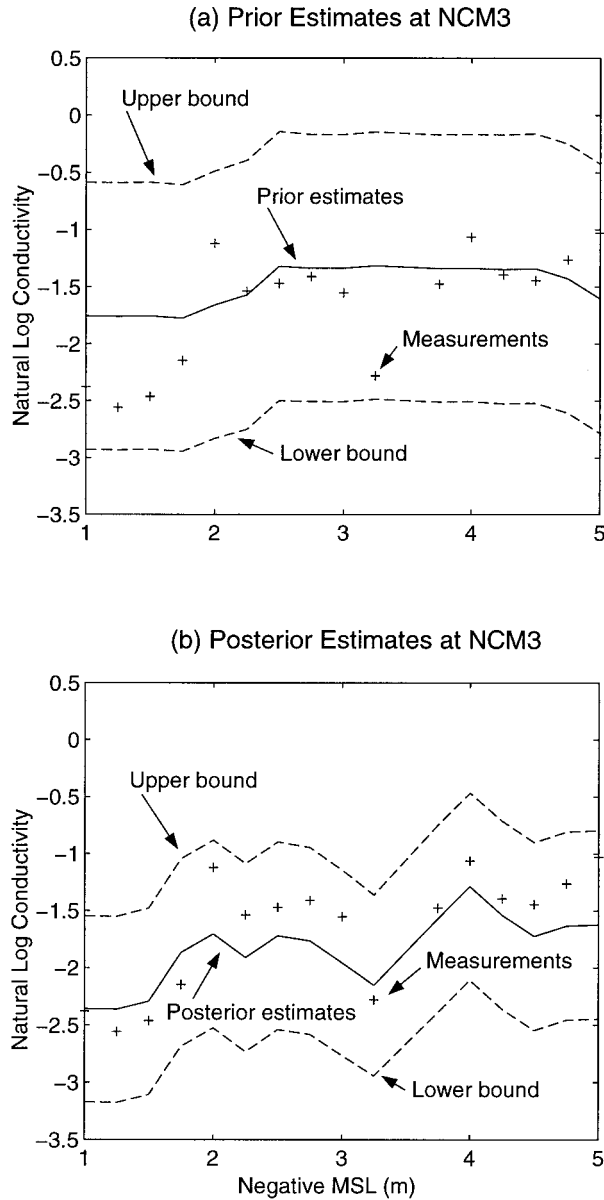

(b) Posterior Estimates at NCM3

Figure 9. The 95% confidence intervals for testing well NCM3.

at a testing well, respectively. Table 2 summarizes the averaged reductions over the 10 testing wells for different models. GPR and seismic velocity were found to be more useful than GPR attenuation as only one type of geophysical data is used. The zero reduction in the standard deviations by GPR attenuation reflects either that the current method does not effectively account for the nonlinear relationship between the log conductivity and the GPR attenuation or that the GPR attenuation is noninformative at this site. As two or more types of geophysical data are used, the average reductions in the standard deviations increase but only incrementally. This is the outcome of the various degrees of correlations that exist between the various types of data which leads to redundancy in data and hence to only minor improvements.

5. Discussion and Conclusions

Our paper explored the use of geophysical tomographic data for hydraulic conductivity estimation using a Bayesian frame-
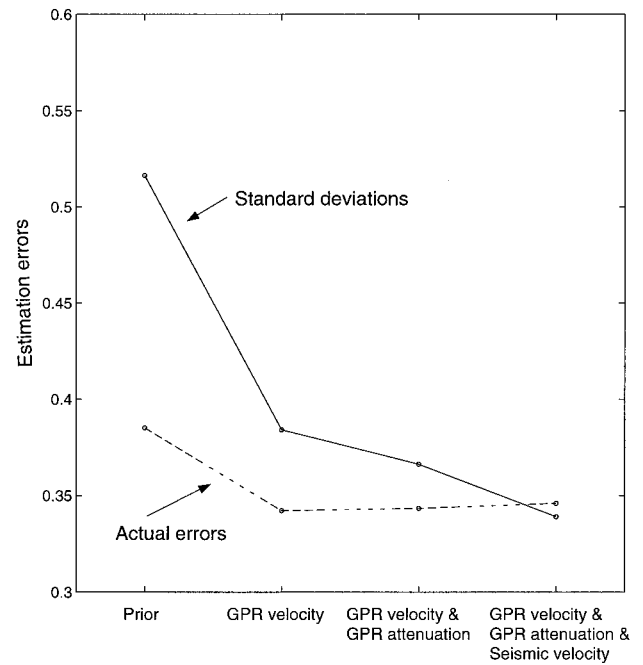

Figure 10. Comparison of the standard deviations and actual errors, which are the space averages along the wells of the absolute differences between actual and estimated values.

work. The prior estimates were inferred from the hydraulic conductivity data measured at well bores, and the posterior estimates were obtained by updating the prior using collocated geophysical data. Both the prior and the posterior estimates were compared with the actual measurements to evaluate the usefulness of geophysical data for hydraulic conductivity estimation. The key findings of this study are summarized below.

Geophysical tomographic data hold the potential to improve estimation of hydraulic conductivity even when log conductivity displays small variations and geophysical data vary over narrow ranges. This is possibly true also for domains of large variability, since large variability implies large contrasts in the geophysical measurements and, consequently, better correlations between the hydraulic conductivity and the geophysical measurements. Thus the advantages of using geophysical data in the hydraulic conductivity estimation may become more evident, as shown by *Rubin et al.* [1992], *Copty et al.* [1993], *Hyndman et al.* [1994], *Hubbard et al.* [1997], *Ezzedine et al.* [1999], and *Hubbard and Rubin* [2000].

The Bayesian approach coupled with the normal linear re-

Table 2. Reductions of the Standard Deviations Using Various Types of Geophysical Data

Number of Data Types	Names of Data Types	Average Reduction, %	Standard Errors, %
1	v_g	25.3	3.5
	α	0.0	N/A
	v_s	29.2	3.5
2	v_g, α	28.7	3.6
	v_g, v_s	31.9	2.7
	α, v_s	31.0	4.1
3	v_g, α, v_s	34.0	4.1

v_g, GPR velocity; v_s, seismic velocity; and α, GPR attenuation.

gression model is effective in combining geophysical data into hydraulic conductivity estimation. Petrophysical models relating hydraulic conductivity to geophysical measurements are often nonlinear and hard to define [*Hyndman et al.*, 1994; *Ezzedine et al.*, 1999], yet well-defined models are critical for successfully employing geophysical data in the estimation. In our study, petrophysical models are summarized in the form of likelihood functions, and each of the likelihood function is expressed as the product of several conditional pdfs (equation (2)), which were defined using the normal linear model. This method alleviates the difficulty commonly encountered in the inference of petrophysical models for multivariate dependent variables by following a systematic model selection procedure. Another advantage of the method is that the normal linear model provides flexibility in fitting the nonlinear relations between hydraulic conductivity and geophysical measurements since the likelihood functions can be of an arbitrary shape even if each of the conditional pdfs (equation (2)) are assumed to be normal. However, this method is limited in situations where each of the conditional pdfs (equation (2)) is multimodal and asymmetrical. In this case, we can use other techniques, such as described by *Ezzedine et al.* [1999] and *Hubbard and Rubin* [2000].

References

Ahmed, S., G. De Marsily, and A. Talbot, Combined use of hydraulic and electrical properties of an aquifer in a geostatistical estimation of transmissivity, *Ground Water*, 26(1), 78–86, 1988.

Bachrach, R., J. Dvorkin, and A. Nur, High-resolution shallow-seismic experiments in sand, part II, Velocities in shallow unconsolidated sand, *Geophysics*, 63(4), 1234–1240, 1998.

Bachrach, R., J. Dvorkin, and A. Nur, Seismic velocities and poisson's ratio of shallow unconsolidated sands, *Geophysics*, 65(2), 559–564, 2000.

Bernardo, J. M., and A. F. Smith, *Bayesian Theory*, John Wiley, New York, 1994.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley-Longman, Reeding, Mass., 1973.

Carman, P. C., *Flow of Gases Through Porous Media*, Academic, San Diego, Calif., 1956.

Copty, N., and Y. Rubin, A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resour. Res.*, 31(7), 1673–1686, 1995.

Copty, N., Y. Rubin, and G. Mavko, Geophysical-hydrological identification of field permeabilities through bayesian updating, *Water Resour. Res.*, 29(8), 2813–2825, 1993.

Dagan, G., Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 2, The solute transport, *Water Resour. Res.*, 18(4), 835–848, 1982.

Davis, J., and A. Annan, Ground penetrating radar for high resolution mapping of soil and rock stratigraphy, *Geophys. Prospect*, 37, 531–551, 1989.

Deutsch, C., and A. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York, 1998.

Ezzedine, S., Y. Rubin, and J. Chen, Hydrological-geophysical bayesian method for subsurface site characterization: Theory and application to llnl superfund site, *Water Resour. Res.*, 35(9), 2671–2683, 1999.

Gelhar, L. W., and C. L. Axness, Three-dimensional stochastic analysis

of macrodispersion in aquifer, *Water Resour. Res.*, 19(1), 161–180, 1983.

Hubbard, S., and Y. Rubin, Hydrogeological parameter estimation using geological data: A review of selected techniques, *J. Contam. Hydrol.*, 45, 3–34, 2000.

Hubbard, S., Y. Rubin, and E. Majer, Ground penetrating radar assisted saturation and permeability estimation in bimodal systems, *Water Resour. Res.*, 33(5), 971–990, 1997.

Hubbard, S., Y. Rubin, and E. Majer, Spatial correlation structure using geophysical and hydrogeological data, *Water Resour. Res.*, 35(6), 1809–1825, 1999.

Hyndman, D. W., J. M. Harris, and S. M. Gorelick, Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, 30(7), 1965–1977, 1994.

Isaaks, E. H., and R. M. Srivastava, *Applied Geostatistics*, Oxford Univ. Press, New York, 1989.

Johnson, W. P., et al., Ferrographic tracking of bacterial transport in the field at the Narrow Channel Focus Area, Oyster Site, Va., *Environ. Sci. Technol.*, 35, 182–191, 2001.

Journel, A. G. (Ed.), *Fundamentals of Geostatistics in Five Lessons*, Short Course Geol. Ser., vol. 8, AGU, Washington, D. C., 1989.

Journel, A. G., and C. J. Huijbregts, *Mining Geostatistics*, Academic, San Diego, Calif., 1978.

Keers, H., L. R. Johnson, and D. W. Vasco, Acoustic crosswell imaging using asymptotic waveforms, *Geophysics*, 65(5), 1569–1582, 2000.

Kelly, W., Geoelectrical sounding for estimating hydraulic conductivity, *Ground Water*, 15(6), 420–425, 1977.

Kitanidis, P. K., Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, 22(4), 499–507, 1986.

Kitanidis, P. K., Orthonomal residuals in geostatistics: Model criticism and parameter estimation, *Math. Geol.*, 23(5), 741–758, 1991.

Knoll, M., A Petrophysical Basis for Ground Penetrating Radar and Very Early Time Electromagnetics: Electrical Properties of Sand-Clay Mixtures, Ph.D. thesis, Univ. of B. C., Can., 1996.

Marion, D., Acoustical, Mechanical and Transport Properties of Sediments and Granular Materials, Ph.D. thesis, Stanford Univ., Stanford, Calif., 1990.

Marko, G., T. Mukerji, and J. Dvorkin, *The Rock Physics Handbook Tools for Seismic Analysis in Porous Media*, Cambridge Univ. Press, New York, 1998.

Mixon, R. B., Stratigraphic and geologic framework of the upper cenozoic deposits in the southern delmarva peninsula, *U.S. Geol. Surv. Prof. Pap.*, 1067G, 53 pp., 1985.

Molz, F., and S. Young, Development and application of borehole flowmeters for environmental assessment, *The Log Anal.*, 3, 13–23, 1993.

Nihei, K. T., Micromechanics of seismic wave propagation in granular rocks, Ph.D. thesis, Univ. of Calif., Berkeley, 1992.

Parker, R. L., *Geophysical Inverse Theory*, Princeton Univ. Press, Princeton, N. J., 1994.

Peterson, J. E., Pre-inversion processing and analysis of tomographic radar data, *J. Environ. Eng. Geophys.*, in press, 2001.

Peterson, J. E., B. N. Paulsson, and T. V. McEvilly, Applications of algebraic reconstruction techniques to crosshole seismic data, *Geophysics*, 50, 1566–1580, 1985.

Rector, J. W., Crosswell methods: Where are we, where are we going?, *Geophysics*, 60(6), 629–630, 1995.

Rubin, Y., G. Mavko, and J. Harris, Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, *Water Resour. Res.*, 28(7), 1809–1816, 1992.

Rubin, Y., S. Hubbard, A. Wilson, and M. Cushey, Aquifer characterization, in *The Handbook of Groundwater Engineering*, edited by J. W. Delleur, CRC Press, Boca Raton, Fla., 1998.

Scheibe, T. D., Y. Chien, T. Ginn, C. Murray, and S. S. Hubbard, Heterogeneity characterization for field-scale bacterial transport modeling, *Eos Trans. AGU*, 80(17), Spring Meet. Suppl., S115, 1999.

Shapiro, S. A., P. Audigane, and J.-J. Royer, Large-scale in site permeability tensor of rocks from induced microseismicity, *Geophys. J. Int.*, 137, 207–213, 1999.

Stone, C. J., *A Course in Probability and Statistics*, Duxbury, Boston, Mass., 1995.

Urish, D. W., Electrical resistivity-hydraulic conductivity relationships in glacial outwash aquifers, *Water Resour. Res.*, 17(5), 1401–1408, 1981.

Vasco, D. W., J. E. Peterson, and E. L. Major, A simultaneous inver-

sion of seismic traveltimes and amplitudes for velocity and attenuation, *Geophyics*, *61*(6), 1738–1757, 1996.

Waldrop, W., and J. Hamby, Results of field tests with the elecromagnetic borehole flowmeter at the narrow channel focus area, tech. rep., Quantum Eng. Corp., Loudon, Tenn., 1999.

Welhan, J., and M. Reed, Geological analysis of regional hydraulic conductivity variations in the snake river plain aquifer, eastern Idaho, *Geol. Soc. Am.*, *109*(7), 855–868, 1997.

Williamson, P. R., and M. H. Worthington, Resolution limits in ray tomography due to wave behavior: Numerical experiments, *Geophysics*, *58*(5), 727–735, 1993.

J. Chen and Y. Rubin, Department of Civil and Environmental Engineering, University of California, 440 Davis Hall, Berkeley, CA 94720. (rubin@ce.berkeley.edu)

S. Hubbard, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.